UNITED
NATIONS

E

Economic and Social Council

Distr.
LIMITED

E/CONF.90/L.12
11 April 1997

ENGLISH ONLY

SIXTH UNITED NATIONS REGIONAL
CARTOGRAPHIC CONFERENCE FOR
THE AMERICAS
New York, 2-6 June 1997
Item 5 (a) of the provisional agenda*

REVIEW OF THE LATEST TECHNOLOGY IN CARTOGRAPHIC DATA ACQUISITION, MANIPULATION, STORAGE AND PRESENTATION, WITH SPECIAL EMPHASIS ON POTENTIAL APPLICATIONS IN DEVELOPING COUNTRIES: AUTOMATED MAPPING PROJECTS: DEVELOPMENT AND APPLICATION OF DIGITAL CARTOGRAPHIC DATABASES, INCLUDING DIGITAL TERRAIN MODELLING

The Inter-American Geospatial Data Network: developing a western hemisphere geospatial data clearinghouse

Paper submitted by the United States of America**

* E/CONF.90/1.

** Prepared by Michelle L. Anthony, Jacqueline M. Klaver and Robert J. Quenzer, Hughes STX Corporation, EROS Data Center, Sioux Falls, South Dakota (work performed under a United States Geological Survey contract).

97-09963 (E) 300497 050597

INTRODUCTION

Establishing the Inter-American Geospatial Data Network (IGDN) is a project of the U.S. Agency for International Development (USAID) and the U.S. Geological Survey (USGS), in cooperation with organizations throughout the Americas. The objective is to promote access to information that describes the existence and availability of geospatial data sets by means of the Internet. IGDN participants are producing digital geographic information and applying it to problems of sustainable development. The project is being carried out in the context of USAID's efforts to promote environmentally sound economic development. In its efforts to strengthen the Western Hemisphere's information infrastructure, it is consistent with the goals of the Plan of Action that emerged from the 1994 Summit of the Americas.

Geographic information system (GIS) technology is being used increasingly by professionals in a wide variety of disciplines to plan and implement scientific research, commercial investments, environmental policy, civil infrastructure management, and conservation activities. Consequently, there is a growing demand for digital geographic data sets, especially in base categories such as topography, hydrography, roads, populated places, and administrative boundaries. The IGDN is envisioned as an online clearinghouse to facilitate the discovery of existing data sets and to avert duplication of effort in creating new data sets. Two tasks central to the IGDN are the creation of standardized, descriptive text files, or "metadata," and the use of search engine technology to support remote query and retrieval.

CLEARINGHOUSE TECHNOLOGY

The IGDN is based on the same technology used by the Federal Geographic Data Committee (FGDC) to establish the U. S. National Geospatial Data Clearinghouse. In 1994, President Clinton signed U.S. Executive Order 12906, "Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure" (White House, 1994). The order directs U.S. agencies to document geospatial data and make them available to the public. It calls upon the FGDC to develop the standards, organizational guidelines, and software solutions that assist organizations as they work to implement the Executive Order. A major activity of the FGDC is the development of a national clearinghouse as a distributed, electronically connected network of geospatial data producers, managers, and users. A clearinghouse allows its users to determine what geospatial data exist, evaluate the usefulness of the data, and obtain data.

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

The USGS is using FGDC standards and guidelines to establish the IGDN as a geospatial data clearinghouse for the Western Hemisphere. The IGDN provides a consistent data discovery mechanism that will help minimize geospatial data duplication, promote data availability, and coordinate data collection and research activities.

An IGDN node houses and serves an organization's metadata and optionally its geospatial data. Initially, five organizations have been invited to participate from government, academic, nongovernment organizations, and the commercial sector. These initial nodes include CODEPLAN, the planning agency for the Federal District of Brasilia, Brazil; Conservation International, an environmental nongovernment organization in Washington, D.C.; GEOSCAPE International, a commercial firm in Boca Raton, Florida; IBGE, the Brazilian mapping and statistical agency in Rio de Janeiro, Brazil; and the University of Kansas GIS and Environmental Modeling (GEM) Laboratory in Lawrence, Kansas. The IGDN membership is being further expanded to include potential participation in the geospatial data community through USAID Missions, current IGDN members, the Pan American Institute of Geography and History (PAIGH), and other professional organizations.

METADATA

Metadata are at the heart of the IGDN. Metadata consist of standardized text files describing the content, quality, constraints, availability, format, means of access, and cost of existing geospatial data sets. Major uses of metadata include the ability to represent an organization's data holdings to clearinghouses and provide descriptive information necessary to process and interpret geospatial data sets.

Geospatial data can be difficult or impossible to use if unaccompanied by metadata. Metadata help users determine the suitability of geospatial data for their specific purpose. Data origin, accuracy, and attributes are fundamental elements in determining if the data are appropriate for a users analytical or cartographic needs. Data constraints, availability, format, means of access, and cost are elements used in evaluating whether data can be successfully integrated with an existing or planned GIS.

Metadata preparation and maintenance present a major challenge to organizations developing an IGDN node. Organizations need to develop a plan for documenting both new and existing geospatial data. An inventory of existing geospatial data is a first step to identify an organization's geospatial data holdings, as well as their location, condition, and content. The inventory will help define which data to document and include in the IGDN. Strategies and priorities for documenting the data need to be established on the basis of the data's greatest potential value to users. Criteria that may be used are the number and identity of interested users, unique nature of the data, age of the data, size of the data set, and cost of documenting the data. Trained personnel are essential to maintain current, high-quality metadata.

The rising popularity of the Internet has significantly increased access to georeferenced data. Because overwhelming amounts of information are available to the public, methods have been developed for geospatial data search and retrieval. A metadata content standard has been developed to provide the structure required for an effective geospatial data search on the Internet. The IGDN adheres to the metadata standard established by the FGDC, described in the "Content Standards for Digital Geospatial Metadata Workbook" Version 1.0 (FGDC, 1995). The standard defines a common set of approximately 300 terms and definitions to help an organization document its geospatial data. It does not, however, tell a producer how a computer system should prepare the information, transfer metadata and geospatial data, or present information to the user. The FGDC metadata content standard is the basis for a proposed international standard currently being reviewed by the International Organization for Standardization (ISO).

SEARCH AND RETRIEVAL

There are two levels of participation in the IGDN. Table 1 explains each level and its characteristics. Level 1 participants operate established, fully functioning IGDN nodes. Level 2 participants are organizations that have stated interest in pursuing node development with the IGDN but have not yet completed the preparation of metadata, installation of the search engine software, or development of the Internet interface.

The IGDN Clearinghouse uses a client/server model of computing to connect the numerous data providers and users on the Internet. The IGDN client consists of an Internet browser and a geospatial search form. The client sends the user's query to the server and receives the results. The server is an IGDN node consisting of four components: metadata describing geospatial data, hardware for storing indexed metadata, a dedicated Internet connection, and search and retrieval software. The server performs the query on the metadata and sends the query results to the client.

A search of the data is performed on the available IGDN nodes using the Internet and software that is compliant with the Z39.50 protocol. The Z39.50 protocol defines the way that two computers interact with each other. This open standard ensures interoperability of multiple machines and platforms. It has been approved by both the National Information Standards Organization (NISO) in 1995 and the ISO in 1991.

The searcher does not have to be familiar with details of the Z39.50 protocol to make an IGDN query. The searcher uses a browser to fill in the IGDN search form by specifying the IGDN nodes to be queried, a geographic area of interest, and thematic keywords. After the form is submitted, the client converts the query into a Z39.50 standardized format and sends it to the selected IGDN nodes over the Internet.

Each selected IGDN node receives the same query for processing. The search and retrieval software simultaneously processes the queries against the FGDC metadata on each node, significantly reducing the time required to return results. Nodes with matching metadata return brief descriptions of the data displayed on the user's browser.

Each metadata description contains associated hyperlinks to the full metadata files. The hyperlink allows the user to view the metadata in the browser. A user can then evaluate information in the metadata to determine if the data meet his or her specific needs.

The Z39.50 search and retrieval software allows the IGDN to be a decentralized system of data producers. The software is multi-threaded, allowing many servers to be searched at the same time. It also permits multiple users to conduct queries with each of the IGDN nodes simultaneously. The Z39.50 software allows the IGDN to be a distributed system of geospatial data nodes that document their data in the same vocabulary and support the same protocol.

IGDN STATUS

The IGDN project was established in October 1995. Project staff have developed a prototype web site that includes a fully functional search engine connected to IGDN nodes, provided participating organizations with technical assistance to establish their nodes, and developed outreach programs to increase the number of future participants. The IGDN can be accessed on the Internet at the Uniform Resource Locater (URL) *http://edcintl.cr.usgs.gov/igdn/igdn.html*.

Approximately 10 organizations have expressed interest in IGDN participation at level 1 or 2. Level 1 participants will continue to develop and maintain their IGDN nodes. Level 2 participants will receive technical assistance from the USGS to develop IGDN implementation strategies. Future efforts will be made to expand the number of Western Hemisphere geospatial data organizations participating in the project.

A workshop hosted by USAID, the USGS, and PAIGH will be held in July 1997. The workshop will train new IGDN participants in metadata preparation and standards, network searching software, data transmission and delivery, and web development. Scientists from participating organizations will learn to implement the IGDN technology as well as train new IGDN participants within their respective countries.

REFERENCES

White House (1994) Executive Order 12906, Coordinating Geographic Data Acquisition and Access: the National Spatial Data Infrastructure. Washington, D.C., April 11, 1994.

Federal Geographic Data Committee (1995) *Content Standards for Digital Geospatial Metadata Workbook (March 24)*, Federal Geographic Data Committee, Washington, D.C.

Table 1. Levels of IGDN Participation

LEVEL & NAME	DESCRIPTION	REQUIREMENTS / IMPLEMENTATION
Level 1 IGDN Nodes	Full Federal Geographic Data Committee (FGDC) Clearinghouse implementation	• Western Hemisphere geospatial data • Metadata in FGDC format • Dedicated Internet connection • Search & retrieval software on UNIX or Windows NT • Organization's node is connected to the IGDN search engine's node registry • Organization listed as IGDN Node Page
Level 2 Future IGDN Nodes	Organizations interested in future clearinghouse node implementation	• Organization is a holder/distributor of geospatial data and has a WWW site • Contact has been established with IGDN staff to discuss project objectives • Organization concurs with IGDN concept and will develop plans for future integration • Organization's name, logo, and WWW site listed on Future IGDN Node Page

---